# Gaming and Live Streaming Commerce Platform

**Problem**  
The platform needed to serve three very different audiences at once.  
Viewers wanted the right live games and sports, plus one-tap purchases inside the stream.  
Influencers wanted growth, fair monetization, and simple tools to showcase products while going live.  
Vendors wanted a clean way to list products, match with the right influencers, negotiate deals, and see what actually converts.  
Doing all of that at scale required a brain behind the scenes, not just another streaming app.

**Solution**  
We built a streaming-first marketplace with AI at the core.

Viewer experience  
Recommendation models using collaborative filtering and sequence models ranked streams, highlights, and VODs per user. Real-time embeddings updated interests from watch time, chat activity, and purchase signals. A live-ranking model re-ordered the homepage as events spiked, so the best streams surfaced while they were hot.

Influencer tooling  
A creator growth engine predicted audience segments likely to convert during upcoming shows using gradient boosting and time-series features. A content quality scorer used vision and audio models to flag low bitrate, bad lighting, or clipping so creators could fix issues before going live. An in-stream offer selector suggested which SKUs to pin based on current chat sentiment and historical uplift.

Vendor and commerce  
A vendor–influencer matching model used two-tower retrieval with product, audience, and creator embeddings to propose partnerships. A negotiation copilot used an instruction-tuned LLM to draft first offers, counteroffers, and contract clauses with clear guardrails. Dynamic pricing and promotion models tested price elasticity on limited cohorts before rolling out to everyone.

Trust, safety, and support  
Multimodal moderation combined text and vision models to filter hate, adult content, and IP violations in real time. A payments risk model with graph features detected collusive refunds and synthetic purchases. An LLM support agent resolved common issues instantly and escalated edge cases with full context.

**Process**

1. Data and events  
   We set up real-time ingestion for views, likes, chat, purchases, and stream health. Features landed in an online feature store to keep training and inference in sync.
2. Personalization  
   We trained collaborative filtering baselines, then moved to deep sequence recommenders that learned watch and buy patterns over sessions. Cold-start users were handled with content-based models on game metadata, tags, and language.
3. Matching and negotiation  
   Two-tower retrieval learned vendor and influencer vectors in the same space. A re-ranker added margin, audience overlap, and compliance. The negotiation assistant was fine-tuned on anonymized deal threads with templates aligned to legal policy.
4. Commerce and pricing  
   Uplift models estimated the incremental impact of placing a product at a specific moment in a stream. Dynamic pricing experiments ran safely with guardrails for vendor floors and regional regulations.
5. Safety and quality  
   Text and image models ran pre-chat and mid-stream. A QoS optimizer adjusted encoder settings and CDN routes based on predicted churn if the bitrate dipped.
6. Measurement  
   Attribution used sequence-aware multi-touch models to credit creators and vendors fairly. We shipped an experimentation framework for rapid A/Bs across the homepage, pin-to-buy moments, and checkout UI.

**Models and techniques used**  
Recommendations and ranking  
Collaborative filtering for baselines  
Deep sequence models for next-stream and next-purchase prediction  
Real-time embeddings with approximate nearest neighbor retrieval  
Contextual bandits to explore new creators and offers safely

Creator growth and content quality  
Gradient boosting for show-level conversion forecasts  
Vision and audio classifiers for stream health scoring

Vendor–influencer marketplace  
Two-tower retrieval for candidate matching, learning-to-rank for final ordering  
Instruction-tuned LLM for negotiation drafts and contract language suggestions

Commerce and pricing  
Uplift modeling for placement impact  
Dynamic pricing with constrained optimization  
Vector search for product discovery in chat

Trust, safety, and payments  
Multimodal moderation (text and vision)  
Graph-based fraud detection on wallets, cards, and devices  
LLM support assistant with tools for refunds, subscription changes, and KYC checks

Platform and infra notes  
Event streaming for real-time signals  
Online feature store for low-latency inference  
Microservices for recommendations, matching, pricing, moderation  
Observability with end-to-end tracing from impression to purchase

**Outcome**  
Viewer engagement up 40 percent with personalized stream discovery.  
Influencer revenue per hour up 30 percent through smarter product pinning and audience targeting.  
Vendor deal cycle time cut by half with better matching and negotiation assistance.  
Checkout conversion increased 18 percent on in-stream purchases.  
Policy violations and payment fraud reduced significantly without slowing the experience.

**Impact in one line**  
A single place where fans watch, creators perform, brands sell, and the whole thing feels effortless because the platform is quietly learning in real time.

**Healthcare – AI-Powered Online Consultations**

**Problem**  
Doctors on an online consultation platform were spending too much time on admin.  
Every appointment started with digging through uploaded prescriptions, scanned lab reports, and handwritten notes.  
Information was scattered and unstructured, which meant longer prep, repeated patient histories, and sometimes missed details that mattered.

**Solution**  
We built an AI consultation assistant to prepare doctors automatically.

Document understanding  
OCR models digitized handwritten prescriptions and low-quality scans.  
NLP pipelines (built on transformer models fine-tuned for medical text) extracted symptoms, diagnoses, and medications from unstructured documents.  
Contextual summarizers condensed patient history into concise, doctor-ready notes before each consultation.

Workflow integration  
Summaries and extracted data were pushed directly into the online consultation platform, available as soon as a patient booked.  
Doctors saw the whole picture at a glance — past conditions, active prescriptions, and flagged risks — without searching through files.

**Process**

1. Collected anonymized patient documents (scans, PDFs, typed and handwritten notes).
2. Trained OCR on medical prescriptions with domain-specific post-processing for drug names.
3. Fine-tuned ClinicalBERT and other transformer models on structured medical terminology and local languages.
4. Built a summarization pipeline to create short patient histories doctors could trust.
5. Integrated everything into the telemedicine workflow, running securely and in compliance with medical privacy rules.

**Models and techniques used**  
OCR with custom lexicons for medical terms and drug names.  
Transformer-based NLP models (ClinicalBERT, BioBERT) for entity extraction.  
Summarization models tuned to emphasize symptoms, diagnoses, and treatment history.  
Entity linking to map extracted information to standard medical ontologies (ICD codes, RxNorm).

**Outcome**  
Doctors saved 50 percent of pre-consultation prep time.  
Patients no longer had to repeat their medical history during calls.  
Diagnostic accuracy improved as important details were never missed.

**Impact in one line**  
Consultations became faster, smoother, and safer — with doctors spending time on care, not paperwork.

**Automobile – Predictive Maintenance for Fleets**

**Problem**  
A fleet operator managing hundreds of vehicles faced costly breakdowns and unplanned downtime.  
Preventive maintenance based on mileage or time wasn’t enough — engines, brakes, and electronics still failed unpredictably.  
Each breakdown meant revenue loss, emergency repairs, and unhappy customers.

**Solution**  
We built a predictive maintenance system to keep vehicles on the road.

Real-time data capture  
IoT sensors embedded in vehicles tracked engine health, fuel efficiency, temperature, and braking patterns.  
Telemetry was streamed continuously into a central system.

AI-powered failure prediction  
Time-series deep learning models (LSTMs) learned normal vs failing performance patterns.  
Anomaly detection flagged deviations invisible to traditional monitoring.  
Risk scores were generated per vehicle to prioritize inspections.

Fleet dashboards  
Fleet managers received dashboards showing which vehicles needed attention, why they were at risk, and recommended actions.

**Process**

1. Deployed IoT devices across fleet vehicles to collect engine and subsystem metrics.
2. Streamed telemetry into a cloud pipeline for preprocessing and feature extraction.
3. Trained LSTM models on historical “healthy” vs “failure” sequences.
4. Layered anomaly detection to catch edge cases and outliers.
5. Built real-time dashboards and alert systems for fleet managers and service teams.

**Models and techniques used**  
IoT data ingestion and preprocessing for noisy sensor signals.  
Time-series forecasting with LSTMs for degradation trends.  
Anomaly detection using isolation forests and autoencoders.  
Predictive risk scoring combining multiple signals into a single maintenance index.

**Outcome**  
Breakdowns reduced by 40 percent across the fleet.  
Warranty and repair costs fell significantly.  
Vehicle uptime improved, leading to more reliable service and happier customers.

**Impact in one line**  
Vehicles became proactive about their own health, preventing failures before they happe

**Virtual Avatars – Human Interaction at Scale**

**Problem**  
Digital platforms across industries were facing the same challenge: how to provide human-like interaction to millions of users without scaling staff at the same rate.

* In healthcare, patients wanted real-time guidance before consultations.
* In retail, shoppers wanted personalized recommendations from an assistant that felt human.
* In education and entertainment, users wanted engagement, not just scripted bots.

Traditional chatbots couldn’t deliver. They lacked personality, emotional nuance, and contextual understanding. Most importantly, they didn’t feel alive.

The platform needed a way to create digital avatars that could see, listen, and respond like people — available 24/7, at scale, and adaptable to different domains.

**Solution**  
We built an AI-powered avatar system that combined conversational intelligence, visual realism, and multimodal input to create human-like digital assistants.

* **Conversational Intelligence**: NLP and dialogue models powered contextual, natural conversations. Knowledge grounding ensured responses were factually accurate and aligned with domain-specific needs (healthcare, retail, gaming, education).
* **Visual Realism**: Avatars were rendered in 3D with neural rendering techniques. They could mirror facial expressions, gestures, and tone. Lip-sync models ensured speech matched perfectly with animated facial movement.
* **Multimodal Understanding**: Avatars were not limited to text or voice. They could interpret scanned medical documents, uploaded product photos, or even live camera input to guide users.
* **Scalable Deployment**: The system was designed to work seamlessly on web, mobile, and AR/VR platforms, ensuring users experienced consistent quality regardless of device.

The result was an avatar that could act as a doctor’s assistant, a digital shopping guide, a student tutor, or even a live-streaming companion — all powered by the same underlying framework.

**Process**

1. **Speech and Language Pipeline**
   * Built using speech-to-text (STT) models like Whisper and Wav2Vec for fast, multilingual recognition.
   * Dialogue handled by transformer-based NLP models fine-tuned on healthcare, retail, and entertainment datasets.
   * Text-to-speech (TTS) added prosody control so voices sounded natural, with emotion that matched the context.
2. **Avatar Animation**
   * Trained speech-to-animation models to predict visemes (lip movements) aligned with phonemes in real time.
   * Facial animation enhanced with GAN-based rendering for realistic micro-expressions (smiles, frowns, eye blinks).
   * Body gesture models allowed avatars to nod, wave, or emphasize points naturally.
3. **Domain Grounding**
   * Healthcare avatars integrated OCR + NLP to read and summarize medical reports for doctors.
   * Retail avatars connected to product catalogs and recommendation systems, making suggestions as if a human store assistant was guiding.
   * Education avatars integrated knowledge graphs, ensuring subject-specific accuracy in tutoring.
4. **Scalable Infrastructure**
   * Used microservices for speech, rendering, and dialogue so each could scale independently.
   * Deployed real-time streaming with low-latency pipelines to keep lip-sync and conversation aligned.
   * Built AR/VR integrations to bring avatars into immersive environments for live events or metaverse-like experiences.
5. **Continuous Learning**
   * Engagement metrics tracked where users dropped off or re-asked questions.
   * Feedback loops fed into fine-tuning dialogue models, making avatars smarter with real-world use.

**Models and Techniques Used**

* **Speech Recognition**: Whisper, Wav2Vec for real-time multilingual transcription.
* **Dialogue Management**: Transformer-based LLMs fine-tuned for medical, retail, and educational domains.
* **Knowledge Grounding**: Retrieval-augmented generation (RAG) to keep responses factual and domain-specific.
* **Text-to-Speech**: Tacotron and FastSpeech with prosody tuning for expressive, natural voices.
* **Lip-Sync and Animation**: Viseme prediction networks for phoneme-to-face alignment.
* **Facial Realism**: GAN-based neural rendering for emotions and micro-expressions.
* **Multimodal Inputs**: OCR + NLP for documents, vision models for images, gesture recognition for interactive experiences.

**Outcome**

* **Healthcare**: Doctors saved hours per day as avatar assistants prepared patient histories, answered pre-consultation queries, and triaged common symptoms.
* **Retail**: Shoppers engaged longer with digital sales assistants that could recommend, upsell, and guide purchases naturally, boosting conversion rates.
* **Education**: Students reported higher satisfaction and retention with avatars delivering interactive, human-like tutoring compared to static e-learning tools.
* **Entertainment**: Live-streaming avatars served as companions, moderators, and product promoters, creating a hybrid of engagement and commerce.

**Impact in One Line**  
The platform unlocked a new kind of interaction — digital avatars that don’t just respond, but engage, empathize, and guide users as if they were real.